PATENT APPLICATION BASED ON: Docket Number 81961DMW

Inventor(s): Alexander C. Loui
Matthieu Jeanson
Zhaohui Sun

Attorney: David M. Woods

Document ID: \DOCKETS\81961

EVENT CLUSTERING OF IMAGES USING FOREGROUND/BACKGROUND SEGMENTATION

EASTMAN KODAK COMPANY

RESTRICTED INFORMATION

"Express Mail" mailing label number EL486848070US
Date of Deposit March 28, 2001
I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to "Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to the Commissioner of Patents and Trademarks, Washington, D.C. 20231

Robin G. Reeves
(Typed or printed name of person mailing paper of fee)

Robin G. Reeves
(Signature of person mailing paper or fee)

03/28/01

EVENT CLUSTERING OF IMAGES USING FOREGROUND/BACKGROUND SEGMENTATION

FIELD OF THE INVENTION

The invention relates generally to the field of auto albuming of consumer-captured images, and in particular to a system for classifying consumer-captured images by event similarity.

BACKGROUND OF THE INVENTION

Pictorial images are often classified by the particular event, subject or the like for convenience of retrieving, reviewing and albuming of the images. Typically, this has been achieved by manually segmenting the images, or by an automated method that groups the images by color, shape or texture in order to partition the images into groups of similar visual content. It is clear that an accurate determination of content would make the job easier. Although not directed to event classification, there is a body of prior art addressing content-based image retrieval and the content description of images. Some typical references are described below.

In U.S. Patent No. 6,072,904, "Fast image retrieval using multi-scale edge representation of images", a technique for image retrieval uses multi-scale edge characteristics. The target image and each image in the data base are characterized by a vector of edge characteristics within each image. Retrieval is effected by a comparison of the characteristic vectors, rather than a comparison of the images themselves. In U.S. Patent No. 5,911,139, "Visual image database search engine which allows for different schema", a visual information retrieval engine is described for content-based search and retrieval of visual objects. It uses a set of universal primitives to operate on the visual objects, and carries out a heterogeneous comparison to generate a similarity score. U.S. Patent No. 5,852,823, "Image classification and retrieval system using a query-by-example paradigm", teaches a paradigm for image classification and retrieval by query-

by-example. The method generates a semantically based, linguistically searchable, numeric descriptor of a pre-defined group of input images and which is particularly useful in a system for automatically classifying individual images.

The task addressed by the foregoing three patents is one of image retrieval, that is, finding similar images from a database, which is different from the task of event clustering for consumer images, such as photo album organization for consumer images. The descriptors described in these patents do not suggest using foreground and background segmentation for event clustering. Most importantly, the segmentation of images into foreground and background is not taken into account as an image similarity measure.

Commonly-assigned U.S. Patent No. 6,011,595, "Method for segmenting a digital image into a foreground region and a key color region", which issued January 4, 2000 to T. Henderson, K. Spaulding and D. Couwenhoven, teaches image segmentation of a foreground region and a key color backdrop region. The method is used in a "special effects" process for combining a foreground image and a background image. However, the foreground/background separation is not used for image similarity comparison.

Commonly assigned U.S. Patent Application Serial No. 09/163,618, "A method for automatically classifying images into events", filed September 30, 1998 in the names of A. Loui and E. Pavie, and commonly-assigned U.S. Patent Application Serial No. 09/197,363, "A method for automatically comparing content of images for classification into events", filed November 20, 1998 in the names of A. Loui and E. Pavie, represent a continuous effort to build a better system of event clustering for consumer images, albeit with different technical approaches. Serial No. 09/163,618 discloses event clustering using date and time information. Serial No. 09/197,363 discloses a block-based histogram correlation method for image event clustering, which can be used when date and time information is unavailable. It teaches the use of a main subject area (implemented by fixed rectangle segmentation) for comparison, but does not

propose any automatic method of performing foreground/background segmentation, which would be more accurate than a fixed rectangle.

Two articles –one by A. Loui, and A. Savakis, "Automatic image event segmentation and quality screening for albuming applications," *Proceedings IEEE ICME 2000*, New York, Aug. 2000 and the other by John Platt, "AutoAlbum: Clustering digital photographs using probabilistic model merging", *Proceedings IEEE Workshop on Content-based Access of Image and Video Libraries*, 2000 – specifically relate to event clustering of consumer images; however they do not look into regions of images and take advantage of the foreground and background separation. Loui and Savakis teach an event clustering scheme based on date and time information and general image content. Platt teaches a clustering scheme based on probabilistic merging of images. Both of them fail to address the foreground and background separation.

What is needed is a system for segmenting images into coarse regions such as foreground and background and deriving global similarity measures from the similarity between the foreground/background regions. Furthermore, such a system should not become confused by unnecessary details and irrelevant clusters in consumer images.

SUMMARY OF THE INVENTION

The present invention is directed to overcoming one or more of the problems set forth above. Briefly summarized, according to one aspect of the present invention, an event clustering method uses foreground and background segmentation for clustering images from a group into similar events. Initially, each image is divided into a plurality of blocks, thereby providing block-based images. Utilizing a block-by-block comparison, each block-based image is segmented into a plurality of regions comprising at least a foreground and a background. One or more features, such as luminosity, color, position or size, are extracted from the regions and the extracted features are utilized to estimate and compare the similarity of the regions comprising the foreground and background

in successive images in the group. Then, a measure of the total similarity between successive images is computed, thereby providing image distance between successive images, and event clusters are delimited from the image distances.

This invention further includes a system for event clustering of consumer images using foreground/background segmentation, which can be used for auto albuming and related image management and organization tasks. The goal of the disclosed system is to classify multiple consumer photograph rolls into several events based on the image contents, with emphasis on the separation of foreground and background. An important aspect of this system is automatic event clustering based on foreground and background segmentation, leading to better similarity matching between images and performance improvement. Another advantage of the present invention is the use of a block-based approach for segmentation, which will be more computationally efficient than a pixel-based segmentation scheme.

These and other aspects, objects, features and advantages of the present invention will be more clearly understood and appreciated from a review of the following detailed description of the preferred embodiments and appended claims, and by reference to the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a block diagram of event clustering using block-based foreground/background segmentation according to the invention.

Figures 2A and 2B show details of the block-based segmentation technique shown in Figure 1, in particular showing the joining of block boundary separations to form regions.

Figure 3 demonstrates an example of foreground and background segmentation according to the invention.

Figure 4 illustrates the comparison of distance (dissimilarity) measures generated for regions comprising the foreground and background in two images.

Figures 5A, 5B and 5C show the use of memory to compute distance between successive and more distant images in a chronological sequence of such images.

Figure 6 shows an example of foreground and background separation for four consumer images.

Figure 7 shows a similarity comparison between the foreground and background regions of the four images shown in Figure 6.

Figure 8 is a precision recall plot showing the event clustering performance using foreground and background segmentation.

DETAILED DESCRIPTION OF THE INVENTION

In the following description, a preferred embodiment of the present invention will be described in terms that would ordinarily be implemented as a software program. Those skilled in the art will readily recognize that the equivalent of such software may also be constructed in hardware. Because image manipulation algorithms and systems are well known, the present description will be directed in particular to algorithms and systems forming part of, or cooperating more directly with, the system and method in accordance with the present invention. Other aspects of such algorithms and systems, and hardware and/or software for producing and otherwise processing the image signals involved therewith, not specifically shown or described herein, may be selected from such systems, algorithms, components and elements known in the art. Given the system as described according to the invention in the following materials, software not specifically shown or described herein that is useful for implementation of the invention is conventional and within the ordinary skill in such arts.

Still further, as used herein, the computer program may be stored in a computer readable storage medium, which may comprise, for example; magnetic storage media such as a magnetic disk (such as a hard drive or a floppy disk) or magnetic tape; optical storage media such as an optical disc, optical tape, or machine readable bar code; solid state electronic storage devices such as

random access memory (RAM), or read only memory (ROM); or any other physical device or medium employed to store a computer program.

This invention discloses a system for event clustering of consumer images using foreground/background segmentation, which can be used for auto albuming and related image management and organization tasks. It is a challenging task to automatically organize consumer images without any content description into semantically meaningful events. The goal of the disclosed system is to classify multiple consumer photograph rolls into several events based on the image contents, with emphasis on the separation of foreground and background. An important aspect of this disclosure is automatic event clustering based on foreground and background segmentation, leading to better similarity matching between images and performance improvement.

Referring first to Figure 1, an event clustering system according to the invention operates on a group of images 8, which may be images scanned from a roll of film or provided from other sources, such as from a database of images. The images are typically consumer images since that is where the greater value for event clustering may be found, but there is no requirement for the images to be such. The event clustering algorithm is composed of four major modules, as follows:

- A first module 10 for segmenting each of the images in the group into regions comprising a foreground and a background;
- A second module 12 for extracting one or more low-level features, such as luminosity, color, position and size, from the regions comprising the foreground and the background;
- A third module 14 for computing distances (dissimilarities) between successive images considering all the regions in the foreground and the background, meanwhile taking advantage of the memory of frame order; and
- A fourth module 16 for determining the greatest distance between images in the group, including successive images and

more distantly separated images, in order to delimit the clusters.

Since the invention may also be thought of as a method for event clustering, each of the foregoing modules may also be thought of as the steps that would be implemented in performing the method.

Since a fine and accurate segmentation of background and foreground is difficult and computationally expensive, a coarse segmentation of foreground and background is preferred and adequately serves the purpose. Accordingly, in the first module 10, the image is divided into blocks and the dissimilarity between neighboring blocks is computed to connect different block-to-block separations to form regions, as shown in Figures 2A and 2B. More specifically, an image is first divided into rectangular blocks with respect to a grid outline 20. Then, for each rectangular block 22, its distance (dissimilarity) is computed with respect to its neighboring blocks using the features that will be described subsequently in connection with the second module 12. (Preferably, the distances calculated in equations (3) and (4) are used to establish block-to-block dissimilarity.) The greatest distances are then identified and used to establish initial separation boundaries between the rectangular blocks.

Where the initial separation boundaries are isolated from each other or the image border, they are then connected to each other or the image border along intervening block boundaries of greatest remaining distance (as shown by the arrow connections 26 in Figure 2A) until all separation boundaries are connected to form a plurality of regions 28a, 28b ... 28e. Then the regions are merged two by two by computing the distances (dissimilarity) between all the regions 28a ... 28e and merging those regions that have the smallest distances. This is repeated until two combinations of regions remain. Different region characteristics, such as size, position and contact with the image borders, are then used to distinguish background from foreground. For instance, a large centrally positioned combination of regions is likely to be a foreground and the remaining combination of outwardly positioned regions is likely to be a background. As

shown in Figure 2B, this optimally results in two distinct combinations of regions: regions 28a and 28e comprising a background 30 and regions 28b, 28c and 28d comprising a foreground 32. As an example of an actual image, Figure 3 shows the approximate foreground and background segmentation of a lighthouse image using the foregoing block-based approach.

In certain situations, especially where a small region of the image is quite different from the rest of the image, the block-based segmentation process may provide a foreground or a background of only a few blocks. These few blocks may not be sufficient for an accurate background/foreground segmentation. To avoid this outcome, when a predetermined number of regions formed in the segmentation process are each less than a predetermined size, the foreground is approximated by a rectangle of fixed size and position (the predetermined numbers may be empirically determined.) Intuitively, this rectangle position is in the center between left and right borders and just below the center between top and bottom borders. As will be shown later in connection with Figure 8, allowing for this variation from the main segmentation process for these certain situations provides improved results.

While this block-based segmentation is preferred for its simplicity and efficiency, other automated segmentation techniques may be employed. For example, the segmentation method employed in commonly assigned, copending U.S. Patent Application Serial No. 09/223,860, entitled "Method for Automatic Determination of Main Subjects in Photographic Images" filed December 31, 1998 in the names of J. Luo et al., which is incorporated herein by reference, may be used, albeit at a certain price in computational complexity. This segmentation method provides a two-level segmentation, as follows

- A first level composed of several regions, which are homogeneous.
- A second level that groups the regions from the first level to form a foreground, a background and an intermediate region.

In addition, in certain situations the block-based segmentation process may turn up an uncertain region that will best be categorized as an intermediate region since its distance from other regions is not sufficient to clearly associate it with either background or foreground.

After the image has been segmented in the first module 10, one or more low level features such as luminosity, color, position, and size are extracted in the second module 12 from the regions comprising the foreground 30 and the background 32. At this stage, each feature extraction algorithm also has at its disposal the original image information and the mask(s) created as a result of the segmentation, which are used to separate the foreground and background image information. The feature extraction algorithm for luminosity is based on the formula for *YUV* conversion:

$$Y = 0.299 \times R + 0.587 \times G + 0.114 \times B \qquad \text{Eq. (1)}$$

where Y is luminance and RGB represents the color information obtained from individual pixels of the image. The mean luminosity is computed for the regions comprising the foreground and background. The distance between two different regions is simply the absolute value of the difference of these means. Based on this feature, images may be separated into outdoors images, well highlighted images, and images taken during the night, indoor, or in a dark environment.

To compute the color feature of a region, the hue (*H*), intensity (*I*) and saturation (*S*) are first quantized using the equations:

$$\begin{cases} I = \dfrac{R+G+B}{3} \\ S = 1 - \dfrac{\min(R,G,B)}{I} \\ H = \cos^{-1}\left(\dfrac{\frac{1}{2}\times(R-G)+(R-B)}{(((R-G)^2+(R-B)(G-B))^{\frac{1}{2}})} \right) \end{cases} \qquad \text{Eq. (2)}$$

Every region in the image is represented by a color set. To compute the distance between two color sets c_0 and c_1, the distance is calculated and then a component is added to account for the different sizes of the regions, thereby giving more or less emphasis to each component. Given two color set components $m_0=(h_0,i_0,s_0)$ and $m_1=(h_1,i_1,s_1)$, the distance is calculated as follows:

$$d_{m0,m1} = h_{coeff} \times \min(|h_1 - h_0|, h_{\max} - |h_1 - h_0|) + i_{coeff} \times |i_1 - i_0| + s_{coeff} \times |s_1 - s_0|$$

where h_{coeff}, i_{coeff} and s_{coeff} are determined by the user. Eq. (3)

Then distance between the two color sets c_0 and c_1 is determined by

$$d_{c0,c1} = \frac{1}{n_0 \times n_1} \sum_{m_0 \in c_0} \sum_{m_1 \in c_1} c_0[m_0].d_{m_0,m_1}.c_1[m_1] \quad \text{Eq. (4)}$$

where n_0 and n_1 are the number of pixels of regions 0 and 1 and $c[m]$ is the number of pixel in color set c for level m.

It may be further desirable to consider the position and size features of the different regions. For example, higher weights may be assigned to the regions in the central part of the image.

After the low level features and distances have been extracted and the regions comprising the foreground and background have been determined for each image, distances are computed in the module 14 between different regions (resulting from the segmentation) of different images 40 and 42 from the same group, as shown in Figure 4. The goal of this step is to compute the distances between different images, considering all the regions in each image, where the distance metrics are those used for the block-based segmentation (e.g., the luminosity distance and/or the color set distance).

For each image, there are different regions comprising the foreground and background and perhaps further regions comprising an intermediate area. The goal is to compare regions of the same type, e.g.,

foreground to foreground and background to background, except in the case of the intermediate areas, where they are compared with each other and with regions comprising both background and foreground. More specifically, referring to Figure 4, three regions 44a, 44b, 44c comprising the foreground of image 40 are compared to two regions 46a and 46b comprising the foreground of image 42. Likewise, although not separately enumerated, the three regions (indicated by check marks) comprising the background of image 40 are compared to the single region (also indicated by a check mark) comprising the background of image 42. Figure 4 also illustrates the situation of intermediate areas, where the two regions comprising the intermediate areas of images 40 and 42 are compared with each other and with the regions comprising the foreground and background of the two images.

After the distances between the different regions comprising the foreground and background in successive images have been computed, a total distance between the images is computed in module 14 using a harmonic mean equation, as follows:

$$\text{harmonic mean } (a_1, a_2, \ldots, a_n) = \frac{1}{\frac{1}{a_0} + \frac{1}{a_1} + \ldots + \frac{1}{a_n}} \qquad \text{Eq. (5)}$$

where a_i is the dissimilarity (distance) between the individual regions comprising the foreground and background in the respective images.

After the total dissimilarity between successive images has been determined in the module 14, event clusters are determined in module 16 according to the image distance of the respective images. Given the distances between successive images, a threshold may be chosen and all distances above this threshold are determined to be separations between different event clusters. Conversely, differences below the threshold are not to be taken as event separations, and such images belong to the same event. The threshold may be a constant number or a function of the statistical characteristics of the distance

distribution (such as the maximum distance, the average distance, the variance, and so on), or the number of desired clusters, or the entropy of the whole distribution (entropy thresholding is described in N.R. Pal and S.K. Pal. "Entropic Thresholding," *Signal Processing*, 16, pp. 97-108, 1989). In a preferred implementation, the threshold is a function of the average and the maximum distances in the group of images.

Sometimes, there may be a chronological order of several images apparently belonging to the same event, and all are similar except for one (or a few) images in between. To take advantage of the chronological order of the images, memory can be employed not only to compute the distance between successive (that is, adjacent) images, but also to compute the distance between more distantly separated images. As shown in Figures 5A, 5B and 5C, when a decision is made on whether there is an event break, the adjacent images 50 (no memory) may be compared (Figure 5A), every other image 52 (1-image memory) may be compared (Figure 5B) or every other two images 54 (2-image memory) may be compared (Figure 5C). More specifically, the total distance measured by the harmonic mean may be taken between the respective images to determine if the group of images belong to the apparent event.

It facilitates an understanding of the invention to examine an event clustering example for several images using foreground and background separation. Figure 6 shows an example of foreground and background separation for four typical consumer images. Two event breaks are detected, one event break 60 between images 2 and 3, and the other event break 62 between images 3 and 4. The first row of images shows the four images. The second and third rows show the results of 1-level and 2-level foreground and background segmentation. Figure 6 also demonstrates the foreground and background separation results using a block-based approach. The regions comprising the foreground and background between these images are compared for similarity, as shown in Figure 7, and their respective distances are used for event clustering.

A precision recall plot is used to evaluate the event-clustering

algorithm. The recall and precision are defined as

$$\text{recall} = \frac{\#correct + 1}{\#correct + \#missed + 1}$$

$$\text{precision} = \frac{\#correct + 1}{\#correct + \#false\_positive + 1} \qquad \text{Eqs. (6)}$$

where recall indicates how many event breaks are missed and precision shows how many event breaks are falsely detected while there is no event break. The numbers are between 0 and 1. The bigger the numbers, the better the system performance.

The event-clustering algorithm has been tested on 2600 typical consumer images. The recall/precision performance with no memory is shown in Figure 8. The basic approach used the block-based foreground and background separation. The improved approach indicates a combination of block based foreground/background segmentation and, for the special situation described earlier, fixed rectangular foreground/background separation, where the segmentation is simply replaced by a fixed rectangle in the foreground. To this end, the system has achieved precision of 58% and recall of 58% on event clustering of 2600 consumer images using 2-image memory.

The subject matter of the present invention relates to digital image understanding technology, which is understood to mean technology that digitally processes a digital image to recognize and thereby assign useful meaning to human understandable objects, attributes or conditions and then to utilize the results obtained in the further processing of the digital image.

The invention has been described in detail with particular reference to certain preferred embodiments thereof, but it will be understood that variations and modifications can be effected within the spirit and scope of the invention. For instance, the idea of using foreground and background segmentation for event clustering can be extended to using multiple regions as well.

PARTS LIST

10	first module
12	second module
14	third module
16	fourth module
20	grid outline
22	rectangular block
24	initial separation
26	arrow extensions
28	regions
30	background
32	foreground
40	image
42	image
44a...	regions comprising foreground
46a...	regions comprising foreground
50	adjacent image
52	every other image
54	every other two images
60	first event break
62	second event break